

ANALISIS SENTIMEN PUBLIK DAMPAK CORONA TERHADAP EKONOMI INDONESIA MENGGUNAKAN MEDIA SOSIAL TWITTER

Tugas Akhir

**Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat Sarjana
Teknik Informatika**



Dibuat Oleh:

DWIKI SETYAWAN

140708074

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI UNIVERSITAS
ATMA JAYA YOGYAKARTA**

2020

LEMBAR PENGESAHAN

ANALISIS SENTIMEN PUBLIK DAMPAK CORONA TERHADAP EKONOMI INDONESIA MENGGUNAKAN MEDIA SOSIAL TWITTER

Yogyakarta, 16 Januari 2021

DWIKISETYAWAN

140708074

Menyetujui,

Pembimbing I

Pembimbing II

Findra Kartikasari Dewi, ST, M.M, M.T.

Prof Ir Djoko Budiyanto
S, MEng, PhD.

NPP :

01.10.792

NPP :

08.93.463

Penguji I

NPP :

Mengetahui,
Dekan Fakultas Teknologi Industri
Dr.A. Teguh Siswanto
NPP : 09.93.464

Pernyataan Orisinalitas & Publikasi Ilmiah

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Dwiki Setyawan

NPM : 140708074

Program Studi : Teknik Informatika

Fakultas : Teknologi Industri

Judul Penelitian : ANALISIS SENTIMEN PUBLIK DAMPAK CORONA
TERHADAP EKONOMI INDONESIA MENGGUNAKAN MEDIA SOSIAL
TWITTER

Menyatakan dengan ini:

1. Tugas Akhir ini adalah benar tidak merupakan salinan sebagian atau keseluruhan dari karya penelitian lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta atas penelitian ini, berupa Hak untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum atas pelanggaran Hak Cipta dalam pembuatan Tugas Akhir ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 16 Januari 2021

Yang menyatakan,

Dwiki Setyawan

NPM:140708074

Kata Pengantar

Puji dan Syukur kepada Tuhan Yang Maha Esa, atas berkat dan rahmat, kemudahan, petunjuk dan kekuatan yang diberikan-Nya, sehingga penulis dapat menyelesaikan penulisan skripsi yang berjudul “ANALISIS SENTIMEN PUBLIK DAMPAK CORONA TERHADAP EKONOMI INDONESIA MENGGUNAKAN MEDIA SOSIAL TWITTER”, merupakan prasyarat untuk mencapai gelar Sarjana Teknik Informatika. Dalam penyusunan penulisan skripsi ini tidak terlepas dari pihak-pihak yang ikut membantu baik secara langsung maupun secara tidak langsung. Maka dalam kesempatan ini dengan penuh rasa hormat, penulis ingin menyampaikan rasa terima kasih kepada:

1. Allah SWT arena kehendak dan ridha-Nya peneliti dapat menyelesaikan skripsi ini.
2. Universitas Atma Jaya Yogyakarta, yang telah mendidik dan membina penulis sehingga Tugas Akhir ini dapat diselesaikan.
3. Findra Kartikasari Dewi, ST, M.M, M.T. selaku dosen pembimbing I yang selalu memberikan bimbingan arahan yang baik dan memberikan ilmu-ilmu yang baru bagi penulis.
4. Prof Ir Djoko Budiyanto S, MEng, PhD., selaku dosen pembimbing I yang selalu memberikan bimbingan arahan yang baik dan memberikan ilmu-ilmu yang baru bagi penulis.
5. Papah, Mamah, keluarga bogor yang selalu memberikan dukungan dan doa setiap saat.
6. Leo, Alfi, Wailan, Primo teman yang membantu dalam pengerjaan skripsi.
7. Tim Wackanda yang beranggotakan William, julbin, pier yang selalu menemani.
8. The kontrakans yang mendukung dalam pengerjaan skripsi.

ABSTRAK

Virus Corona atau sebutan lainnya adalah Covid-19 (*Corona Virus Disease*) muncul pada bulan November 2019 di Wuhan, China. Seiring berjalannya waktu, virus tersebut sudah tidak lagi dikategorikan dengan sebuah wabah, tetapi dikategorikan pandemi atau sudah menyebar hampir ke seluruh negara di dunia, salah termasuk Indonesia. Munculnya Covid-19 di Indonesia pada bulan februari 2020 sudah mengakibatkan banyak sektor yang dirugikan, tidak hanya pada kesehatan, tetapi juga sektor ekonomi. Hal ini menimbulkan berbagai sentimen dari publik terkait munculnya Covid-19 di Indonesia melalui media sosial Twitter.

Salah satu alasan publik menggunakan media sosial Twitter dikarenakan kemudahan pengguna untuk menyebarkan informasi atau sebuah kejadian secara *real time*. Munculnya banyak sentimen dalam twitter mendorong penulis untuk melakukan penelitian analisis sentimen dengan menggunakan metode *Naïve Bayes Classifier* yang dengan pembandingan yaitu *Support Vector Machine* (SVM) untuk melakukan proses klasifikasi berdasarkan data twitter. Kata kunci untuk dalam menemukan polaritas sentimen adalah berkaitan “ekonomi dengan Covid-19”.

Metode *Naive Bayes Classifier* digunakan untuk model klasifikasi apakah sentimen tersebut merupakan sentimen positif, sentimen negatif, atau sentimen netral. Berdasarkan hasil implementasi dua metode tersebut, metode *Support Vector Machine* (SVM) memiliki akurasi yang lebih tinggi dibanding metode *Naïve Bayes Classifier* yang dimana akurasi sebesar 99,8754% dengan jumlah opini yang lebih condong ke negatif, sedangkan tingkat akurasi dari metode *Naive Bayes Classifier* sebesar 80,3617% dengan condong negatif..

Kata kunci: Covid-19, *SentiStrength*, *Naïve Bayes Classifier*, *Sentiment Analysis*, *Opinion Mining*, *Support Vector Machine* (SVM).

Daftar Isi

| | |
|--|-----------|
| LEMBAR PENGESAHAN | 2 |
| Pernyataan Orisinalitas & Publikasi Ilmiah | 3 |
| Kata Pengantar | 4 |
| ABSTRAK | 5 |
| Daftar Tabel | 9 |
| BAB I..... | 10 |
| PENDAHULUAN | 10 |
| 1.1 Latar Belakang..... | 10 |
| 1.2 Rumusan Masalah | 11 |
| 1.3 Batasan Masalah..... | 12 |
| 1.4 Tujuan Penelitian..... | 12 |
| 1.5 Manfaat..... | 13 |
| 1.6 Metode Penelitian..... | 13 |
| BAB II..... | 14 |
| 2.1 Tinjauan Pustaka..... | 14 |
| BAB III..... | 19 |
| LANDASAN TEORI | 19 |
| 3.3.1 Seleksi | 20 |
| 3.3.2 Preprocessing..... | 20 |
| 3.3.3 StopWords Removal..... | 20 |
| 3.3.4 Stemming..... | 21 |
| 3.3.5 Tokenizing..... | 21 |
| 3.3.6 Case Folding | 21 |
| 3.3.7 Transformasi..... | 21 |
| 3.3.8 Data mining | 22 |
| 3.3.9 Interpretasi dan evaluasi | 22 |
| BAB IV..... | 27 |
| METODOLOGI PENELITIAN | 27 |
| 4.3 Pengumpulan Data..... | 28 |
| 4.4 Preprocessing Data | 31 |
| 4.5 Data Cleansing..... | 31 |
| 4.6 Stopword | 32 |
| 4.7 4.3.2 Klasifikasi..... | 34 |
| 4.8 Akurasi | 34 |
| 4.9 Presisi | 34 |
| 4.10 Recall..... | 35 |
| 4.11 Confusion Matrix..... | 35 |
| BAB V | 38 |
| HASIL DAN PEMBAHASAN | 38 |
| 5.4 Interpretasi data tweet dampak Corona terhadap ekonomi Indonesia | 53 |
| 5.4.1 Topik ekonomi..... | 57 |
| 5.4.2 Topik psbb..... | 57 |
| 5.4.3 Topik PHK..... | 58 |
| 5.4.4 Topik kerugian..... | 59 |

| | | |
|-----------------------------|---|-----------|
| 5.4 | Analisis sentimen terhadap topik pada dampak corona terhadap perekonomian Indonesia | 59 |
| 5.5 | Analisis sentimen terhadap topik ekonomi..... | 60 |
| 5.6 | Analisis sentimen terhadap PSBB | 61 |
| 5.7 | Analisis sentimen terhadap PHK..... | 63 |
| 5.8 | Analisis sentiment terhadap kerugian..... | 65 |
| 5.9 | Hasil analisis dari topik keseluruhan | 66 |
| BAB VI | | 67 |
| KESIMPULAN DAN SARAN | | 67 |



Daftar Gambar

| | |
|--|-----------|
| <i>Gambar 4. 1 Metodologi Penelitian</i> | <i>27</i> |
| <i>Gambar 4. 2 Metodologi Penelitian</i> | <i>29</i> |
| <i>Gambar 4. 3 Pemisahan titik koma dengan spasi.....</i> | <i>30</i> |
| <i>Gambar 4. 4 Hasil penyusunan data manual</i> | <i>31</i> |
| | |
| <i>Gambar 5. 1 Persentase data kelas.....</i> | <i>46</i> |
| <i>Gambar 5. 2 Grafik ROC Negatif K-10 Fold dalam metode Naive Bayes Classifier</i> | <i>47</i> |
| <i>Gambar 5. 3 grafik ROC Neutral K-10 Fold dalam metode Naive Bayes Classifier.....</i> | <i>48</i> |
| <i>Gambar 5. 4 grafik ROC Positif K-10 Fold dalam metode Naive Bayes Classifier</i> | <i>48</i> |
| <i>Gambar 5. 5 grafik ROC Negatif K-10 Fold dalam metode (SVM).....</i> | <i>50</i> |
| <i>Gambar 5. 6 grafik ROC Neutral K-10 Fold dalam metode SVM</i> | <i>50</i> |
| <i>Gambar 5. 7 grafik ROC Positif K-10 Fold dalam metode SVM</i> | <i>51</i> |
| <i>Gambar 5. 8 Perbandingan hasil Metode</i> | <i>52</i> |
| <i>Gambar 5. 9 jumlah tiap kelas sentimen terhadap tweet dampak corona terhadap ekonomi Indonesia 2020</i> | <i>53</i> |
| <i>Gambar 5. 10 Kumpulan Topik yang ramai dibahas terhadap dampak corona pada ekonomi indonesia</i> | <i>54</i> |
| <i>Gambar 5. 11 Hasil sentimen terhadap topik yang dibahas</i> | <i>55</i> |
| <i>Gambar 5. 12 kelas sentimen topik ekonomi</i> | <i>60</i> |
| <i>Gambar 5. 13 analisis sentimen terhadap PSBB</i> | <i>62</i> |
| <i>Gambar 5. 14 Analisis sentimen terhadap PHK.....</i> | <i>64</i> |
| <i>Gambar 5. 15 Analisis sentimen terhadap topik kerugian</i> | <i>65</i> |

Daftar Tabel

| | |
|---|-----------|
| <i>Tabel 5. 1 Pengumpulan data awal.....</i> | <i>38</i> |
| <i>Tabel 5. 2 Pemisahan data</i> | <i>39</i> |
| <i>Tabel 5. 3 Hasil penghapusan kolom pada data</i> | <i>40</i> |
| <i>Tabel 5. 4 Tahap proses case folding</i> | <i>41</i> |
| <i>Tabel 5. 5 Tahap proses Stopword Removal</i> | <i>42</i> |
| <i>Tabel 5. 6 Proses Stemming</i> | <i>43</i> |
| <i>Tabel 5. 7 Hasil pembobotan dengan sentistrength</i> | <i>44</i> |
| <i>Tabel 5. 8 Hasil dari algoritma sentistrength.....</i> | <i>44</i> |
| <i>Tabel 5. 9 Hasil metode Naïve Bayes Classifier</i> | <i>47</i> |
| <i>Tabel 5. 10 hasil metode SVM dengan K-fold 10.....</i> | <i>49</i> |
| <i>Tabel 5. 11 Perbandingan kedua metode.....</i> | <i>52</i> |
| <i>Tabel 5. 12 Dataset pada topik yang dibahas</i> | <i>56</i> |
| <i>Tabel 5. 13 Contoh tweet yang membahas topik ekonomi.....</i> | <i>57</i> |
| <i>Tabel 5. 14 contoh tweet yang membahas topik PSBB</i> | <i>58</i> |
| <i>Tabel 5. 15 tweet terhadap topik PHK.....</i> | <i>59</i> |
| <i>Tabel 5. 16 ttabel tweet yang membahas kerugian.....</i> | <i>59</i> |
| <i>Tabel 5. 17 contoh tweet sentimen negatif topik ekonomi.....</i> | <i>61</i> |
| <i>Tabel 5. 18 contoh tweet sentimen netral pada topik ekonomi.....</i> | <i>61</i> |
| <i>Tabel 5. 19 contoh tweet sentimen positif pada topik ekonomi</i> | <i>61</i> |
| <i>Tabel 5. 20 sentimen negatif public terhadap PSBB.....</i> | <i>62</i> |
| <i>Tabel 5. 21 sentimen netral public terhadap PSBB</i> | <i>63</i> |
| <i>Tabel 5. 22 sentimen positif public terhadap PSBB</i> | <i>63</i> |
| <i>Tabel 5. 23 tweet sentimen negatif terhadap phk</i> | <i>64</i> |
| <i>Tabel 5. 24 tweet sentimen netral terhadap phk</i> | <i>64</i> |
| <i>Tabel 5. 25 tweet sentimen positif terhadap phk.....</i> | <i>65</i> |
| <i>Tabel 5. 26 Contoh sentimen negatif terhadap topik kerugian.....</i> | <i>66</i> |
| <i>Tabel 5. 27 contoh sentimen netral terhadap topik kerugian</i> | <i>66</i> |
| <i>Tabel 5. 28 contoh sentimen positif kerugian</i> | <i>66</i> |

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada 30 Januari 2020 WHO menyatakan bahwa munculnya virus baru yaitu Covid-19 (*Corona Virus Disease*) yang berasal dari kota Wuhan, China. Covid-19 juga sudah tidak dikategorikan sebagai wabah tetapi sudah dikategorikan sebagai pandemi, dikarenakan wabah tersebut sudah masuk hampir ke seluruh negara di dunia [1]. Indonesia termasuk negara yang terdampak virus corona yang menyebabkan kerugian diberbagai sektor. Sektor yang dirugikan tidak hanya sektor kesehatan tetapi juga sektor perekonomian.

Awal muncul Corona di Indonesia pada awal bulan Februari 2020 [2]. Cepatnya penularan virus tersebut akhirnya pada tanggal 29 Februari 2020 pemerintah Indonesia akhirnya mengeluarkan status darurat bencana selama 91 hari. Langkah-langkah untuk menanggulangi bencana tersebut diantaranya yaitu pemerintah mensosialisasikan tentang *social distancing* dimana setiap warga harus menjaga jarak minimal 2 meter terhadap warga lain, melakukan penyuluhan penggunaan masker yang aman, rajin mencuci tangan dengan air mengalir atau menggunakan hand sanitizer serta melakukan karantina mandiri selama 14 hari bagi warga yang kondisinya menunjukkan gejala.

Selain sosialisasi kebiasaan tersebut, pemerintah juga menerapkan pembatasan pelaksanaan kegiatan yang melibatkan banyak warga. Lahirnya peraturan tersebut menyebabkan dibatasinya aktivitas yang melibatkan masyarakat seperti pembukaan pusat perbelanjaan, pertemuan-pertemuan di hotel, penutupan tempat wisata, pembatasan perjalanan dengan menggunakan fasilitas transportasi umum dan fasilitas masyarakat lainnya.

Pembatasan kegiatan yang melibatkan masyarakat tersebut berimbas pada terjadinya penurunan sektor ekonomi sehingga perusahaan-perusahaan tutup, pemangkasan karyawan, keterpurukan ekonomi, dsb. Tentunya kejadian ini menjadi sebuah topik yang kerap dibahas oleh masyarakat melalui beragam media sosial, salah satunya media sosial *twitter*.

Twitter merupakan media sosial *microblog* dimana penggunaanya hanya dapat mengirim pesan, membaca pesan, dan interaksi-interaksi lainnya antar sesama pengguna twitter dengan memiliki kapasitas karakter sebanyak 280 yang sebelumnya layanan media sosial tersebut hanya memiliki 140 karakter [3]. Kemudahan mengekspresikan pendapat, dan memberikan sebuah informasi atau topik membuat masyarakat memilih media sosial *twitter* sebagai wadah untuk menuangkan isi pikiran tentang pendapatnya sebuah isu. Banyaknya pengguna antusias dalam membahas sebuah topik, tentunya banyak muncul sentimen-sentimen dalam interaksinya. Sentimen-sentimen yang muncul beragam, mulai dari muncul sentimen positif, sentimen negatif, atau sentimen netral terhadap sebuah topik.

Penelitian yang dilakukan adalah melakukan analisis sentimen pengguna media sosial *twitter* mengenai dampak corona terhadap ekonomi Indonesia saat ini. Analisis ini memiliki langkah – langkah atau proses, mulai dari proses *text preprocessing* atau pembersihan data dengan menggunakan dataset media sosial *twitter* sampai pada tahap terakhir, yaitu implementasi metode *Naïve Bayes Classifier* (NBC) dan metode *Support Vector Machine* (SVM). Penelitian sentimen ini berfokus pada pengujian akurasi klasifikasi dari metode *Naïve Bayes Classifier* (NBC) dengan metode *Support Vector Machine* (SVM) dan hasil dari dokumen tersebut apakah bersifat positif, negatif, atau netral.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas, terdapat dua rumusan masalah, yaitu sebagai berikut:

- a. Bagaimana sentimen publik terhadap dampak corona terhadap ekonomi Indonesia?
- b. Bagaimana mengklasifikasikan sentimen publik berdasarkan data cuitan pengguna twitter menggunakan algoritma *SentiStrength*?
- c. Bagaimana tingkat keakuratan dari hasil analisis sentimen dengan menggunakan metode *Naïve Bayes*?

1.3 Batasan Masalah

Berdasarkan rumusan masalah diatas, batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian analisis sentimen ini hanya menggunakan data-data media sosial twitter.
2. Data yang diambil terkait dampak corona terhadap ekonomi Indonesia mulai dari tanggal 15 februari 2020 sampai 12 september 2020.
3. Metode *naïve bayes classifier* digunakan sebagai perbandingan.
4. Data cuitan diambil dengan menggunakan Bahasa Indonesia.
5. Pada pengumpulan data pada twitter menggunakan *keyword* atau kata kunci, hastag, mention.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan penelitian ini adalah sebagai berikut:

1. Menganalisis sentimen publik terhadap dampak corona terhadap konomi Indonesia.
2. Mendapatkan hasil dan klasifikasi pendapat publik terhadap dampak corona terhadap ekonomi Indonesia.
3. Mengetahui tingkat keakuratan dari hasil analisis sentimen dengan menggunakan metode *Naïve Bayes Classifier* dengan metode

Support Vector Machine (SVM).

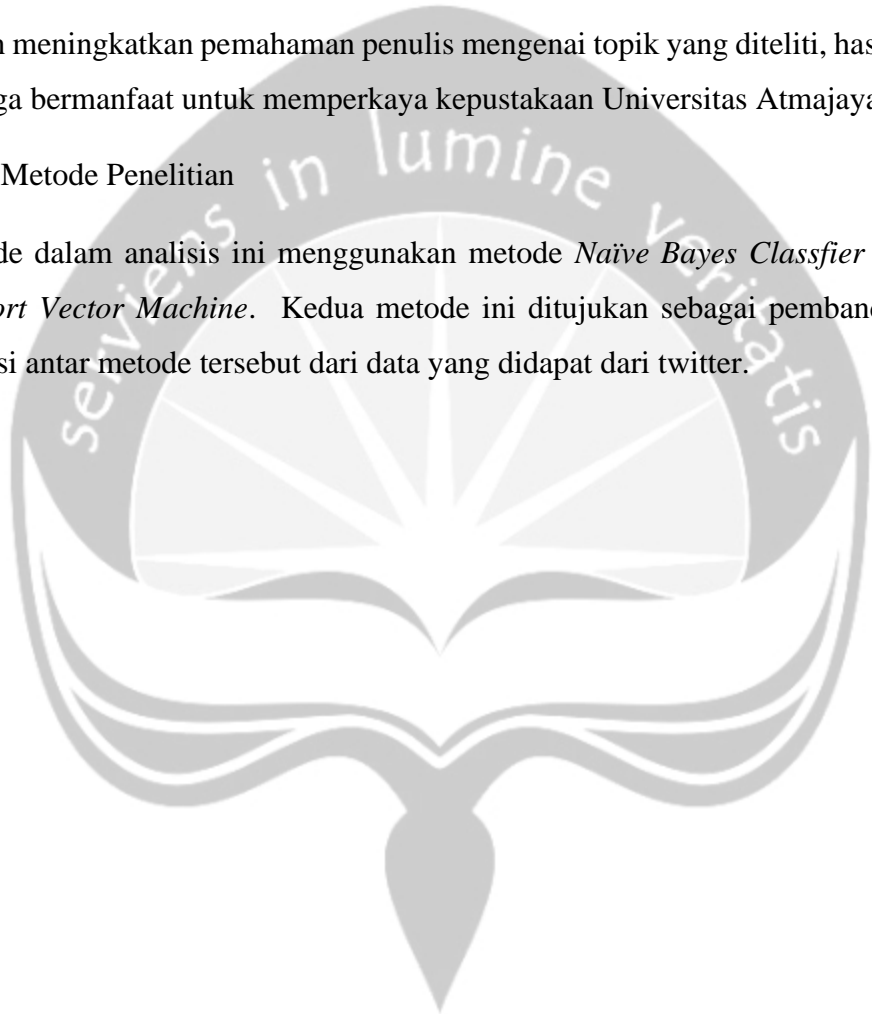
1.5 Manfaat

Manfaat dari penelitian ini mendapatkan hasil dan mengklasifikasikan opini atau pendapat publik berdasarkan media sosial twitter dan diketahuinya tingkat akurasi metoda hasil analisis sentimen dengan menggunakan metoda *Native Bayes*.

Selain meningkatkan pemahaman penulis mengenai topik yang diteliti, hasil penelitian ini juga bermanfaat untuk memperkaya kepustakaan Universitas Atmajaya.

1.6 Metode Penelitian

Metode dalam analisis ini menggunakan metode *Naïve Bayes Classfier* dan metode *Support Vector Machine*. Kedua metode ini ditujukan sebagai pembanding tingkat akurasi antar metode tersebut dari data yang didapat dari twitter.



Berdasarkan hasil dari topik yang dikelompokkan berdasarkan kata kunci yang paling banyak didapat 4 topik kelas memiliki jumlah yang tipis walaupun yang banyak dibahas, diantaranya ekonomi, PSBB, PHK, dan kerugian. Topik ekonomi didominasi dengan sentimen negatif walaupun beda tipis dengan jumlah kelas sentiment lainnya, untuk topik PSBB didominasi dengan kelas sentiment negatif, untuk topik PHK kelas sentiment negatif sangat mendominasi dengan topik ini, dan yang terakhir untuk topik kerugian didominasi dengan sentiment kelas positif. Secara keseluruhan dapat disimpulkan bahwa masing-masing sentiment kelas negatif masih paling banyak dibanding kelas lainnya.



BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Dari hasil penelitian analisis sentimen yang didapat melalui proses studi literatur, pembentukan dataset, penentuan kelas, evaluasi analisis sentimen, sampai tahap terakhir yaitu interpretasi data. Dengan proses yang sudah dilalui,

berikut kesimpulan didapat :

- Dalam analisis ini didapat dataset yang sebanyak 265.307 tweet, algoritma dari Naïve Bayes Classifier (NBC) dengan Support Vector Machine (SVM) menghasilkan akurasi menggunakan metode Naïve Bayes Classifier lebih baik dari pada menggunakan Support Vector Machine dengan algoritma Naïve Bayes Classifier dengan nilai yang diperoleh adalah 86 persen dan menggunakan algoritma Support Vector Machine dengan nilai 99 persen, dari kedua metode tersebut sama-sama menggunakan metode k-10 fold cross validation sehingga tidak perlu menggunakan data training lagi untuk menjalankan berkali-kali.
- Melalui hasil analisis dengan dataset yang diperoleh dan diproses melalui metode NGram pada dataset, didapat berbagai kata yang berhubungan dengan dampak covid terhadap ekonomi Indonesia pada tahun 2020. Setelah melakukan proses pencarian kata kunci terbanyak dari setiap sampel yang didapat, ditemukan topik yang selalu dibahas atau sebuah isu yang ramai dibahas, diantaranya topik ekonomi, topik PSBB, topik PHK, topik kerugian

6.2 Saran

Dari hasil penelitian yang dilakukan tentunya penulis memiliki beberapa kendala dalam melakukan prosesnya, dengan hambatan tersebut diharapkan untuk penulis selanjutnya dapat menyempurnakan penelitian ini. Kendala-kendala tersebut, diantaranya:

- Dalam pencarian data pada proses penelitian ini, penulis masih menggunakan kata kunci dalam kolom pencarian twitter, sehingga topik yang dibahas dalam sampel data tidak selalu pada topik penelitian, dengan anjuran untuk pencarian sampel data menggunakan *hashtag* agar nilai bias dalam penelitian semakin kurang, dan tentunya akurasi dalam penggunaan metode akan meningkat.
- Butuhnya penambahan kosakata dalam kamus SentiStrength, agar penentuan

kelas dalam mengimplementasi algoritma *SentiStrength* semakin akurat pada skor kelasnya.

DAFTAR PUSTAKA

- [1] K. Risnantoyo, ricky ; Nugroho arifin; Mandara, “Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm,” *J. Informatics Telecommunication Eng.*, vol. 4, no. 1, pp. 86–96, 2020.
- [2] A. Karisma, “Dampak Pandemi Corona Terhadap Bisnis Perjalanan dan Wisata di Indonesia,” *SSRN Electron. J.*, no. 120510180016, pp. 1–4, 2020, doi: <http://dx.doi.org/10.2139/ssrn.3590940>.
- [3] D. A. Muthia, “Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes,” *Jurnalilmu Pengetah. Dan Teknol. Komput.*, vol. 2, no. 2, pp. 39–45, 2017.
- [4] P. N. Lhokseumawe, K. Pengantar, rahayu deny danar dan alvi furwanti Alwie, A. B. Prasetyo, and R. Andespa, *Tugas Akhir Tugas Akhir*, vol. 2, no. 1. 2010.
- [5] E. Mas’udah, E. D. Wahyuni, and A. A. Arifiyanti, “Analisis Sentimen: Pemindahan Ibu Kota Indonesia Pada Twitter,” *J. Inform. dan Sist. Inf.*, vol. 1, no. 2, pp. 397–401, 2020.
- [6] R. Nooraeni, H. D. Sariyanti, A. F. F. Iskandar, S. F. Munawwaroh, S. Pertiwi,

- and Y. Ronaldias, “Analisis Sentimen Data Twitter Mengenai Isu RUU KPK Dengan Metode Support Vector Machine (SVM),” *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 1, pp. 55–60, 2020, doi: 10.31294/p.v22i1.6869.
- [7] A. K. Fauziyyah, “Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python,” *J. Ilm. SINUS*, vol. 18, no. 2, p. 31, 2020, doi: 10.30646/sinus.v18i2.491.
- [8] S. S. Salim and J. Mayary, “Analisis Sentimen Pengguna Twitter Terhadap Dompot Elektronik Dengan Metode Lexicon Based Dan K – Nearest Neighbor,” *J. Ilm. Inform. Komput.*, vol. 25, no. 1, pp. 1–17, 2020, doi: 10.35760/ik.2020.v25i1.2411.
- [9] D. Rustiana and N. Rahayu, “Analisis sentimen pasar otomotif mobil:,” *J. SIMETRIS*, vol. 8, no. 1, pp. 113–120, 2017.
- [10] A. Rossi, T. Lestari, R. Setya Perdana, and M. A. Fauzi, “Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Näive Bayes dan Pembobotan Emoji,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1718–1724, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [11] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, “Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 4704–4713, 2018.
- [12] R. Y. Hayuningtyas and R. Sari, “Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata Tmii Menggunakan Näive Bayes Dan Pso,” *J. Techno Nusa Mandiri*, vol. 16, no. 1, pp. 37–42, 2019, doi: 10.33480/techno.v16i1.115.
- [13] D. H. Wahid and A. SN, “Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity,” *IJCCS (Indonesian J.*

- Comput. Cybern. Syst.*, vol. 10, no. 2, p. 207, 2016, doi: 10.22146/ijccs.16625.
- [14] D. J. Haryanto, L. Muflikhah, and M. A. Fauzi, "Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2909–2916, 2018.
- [15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software," *ACM SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009, doi: 10.1145/1656274.1656278.
- [16] C. Sutami, "PERBANDINGAN METODE KLASIFIKASI NAIVE BAYES CLASSIFIER DAN LEXICON BASED DALAM ANALISIS SENTIMEN (Studi Kasus : Twitter)," *J. Chem. Inf. Model.*, 2015, doi: 10.1017/CBO9781107415324.004.
- [17] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel," *J. Khatulistiwa Inform.*, vol. VI, no. 1, pp. 64–69, 2018, [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/3802/2626>.
- [18] J. Bata, Suyoto, and Pranowo, "Leksikon Untuk Deteksi Emosi Dari Teks Bahasa Indonesia," *Semin. Nas. Inform. 2015 (semnasIF 2015)*, vol. 2015, no. November, pp. 195–202, 2015.
- [19] A. Z. Amrullah, A. Sofyan Anas, and M. A. J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *Jurnal*, vol. 2, no. 1, pp. 40–44, 2020, doi: 10.30812/bite.v2i1.804.
- [20] E. H. Muktafin, K. Kusriani, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.

- [21] M. Syarifuddin, "ANALISIS SENTIMEN OPINI PUBLIK TERHADAP EFEK PSBB PADA TWITTER DENGAN ALGORITMA DECISION TREE-KNN-NAÏVE BAYES," vol. 15, no. 1, pp. 87–94, 2020, doi: 10.33480/inti.v15i1.1433.
- [22] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn," *Inti Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020.
- [23] K. Setiawan, B. Rahmatullah, Burhanuddin, A. B. Paryanti, and F. Fauzi, "Komparasi Metode Naive Bayes dan Support Vector Machine Menggunakan Particle Swarm Optimization Untuk Analisis Sentimen Mobil ESEMKA," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 4, no. 3, pp. 102–111, 2020.
- [24] M. M. M. Olhang, "ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP COVID-19 DI INDONESIA MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER(NBC)," 2020, [Online]. Available: <http://eprints.itn.ac.id/id/eprint/5044>.
- [25] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *INTEGER J. Inf. Technol.*, vol. 1, no. 1, pp. 32–41, 2017, [Online]. Available: https://www.researchgate.net/profile/Ghulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf.
- [26] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1725–1732, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628>.

LAMPIRAN

1. Contoh data mentah

| username | text |
|-----------------|---|
| suararaa | Indonesia ga ditutup tutup smp skrng ya pasti ekonomi akan jatuh jg. Mau orang2 ga percaya covid dan sekedar propaganda nurunin ekonomi kalo orng2 pada sakit RS full byk yg meninggal yhaaaa ttp aja nnti pasti menurun ekonominya...";";";"1303803109815197696";https://twitter.com/suararaa/status/1303803109815197696 |
| tempatoceh | Lu lockdown dukung. Lu tau kan efek samping dari lockdown itu apa? Ekonomi indonesia bakal anjlok se anjlok2nya. Terus nanti Jokowi bilang ekonomi kita turun lu marah2 juga. Apa gak tolol kayak gitu?";";";"1303801620170330112";https://twitter.com/tempatoceh/status/1303801620170330112 |
| KontanNews | Indonesia dan Singapura Sepakat Perkuat Kerja Sama Ekonomi http://kabarsuka.com/?p=2803&utm_source=ReviveOldPost&utm_medium=social&utm_campaign=ReviveOldPost â€ |
| clapurhandgirl | Ya... Mereka masih denial terus... Faktanya dulu2x ga mau lockdown. Sekarang dilockdown paksa sama negara lain. Apapun istilahnya. Sekarang Indonesia benar2x ditakuti" oleh negara2x lain. Prestasi luar biasa itu... Selain prestasi pertumbuhan ekonomi Q1 2 97% tentu saja...";";";"1303800090700599297";https://twitter.com/sisgro20/status/1303800090700599297 |
| aokhi | Ya... Silahkan deny aja terus... Faktanya dulu2x ga mau lockdown. Sekarang dilockdown paksa sama negara lain. Apapun istilahnya. Sekarang Indonesia benar2x ditakuti" oleh negara2x lain. Prestasi luar biasa itu... Selain prestasi pertumbuhan ekonomi Q1 2 97% tentu saja...";";";"1303799809602539520";https://twitter.com/sisgro20/status/1303799809602539520 |
| jpnncom | Indef meramal pola pemulihan ekonomi bisa berbentuk L-shaped ini sebabnya http://nasional.kontan.co.id/news/indef-meramal-pola-pemulihan-ekonomi-bisa-berbentuk-l-shaped-ini-sebabnya-1 â€ # EkonomiIndonesia";";";"1303799115608858624";https://twitter.com/KontanNews/status/1303799115608858624 |
| SyahriAlghifari | Ubedilah Badrun: Kejahatan Besar Sedang Terjadi Di Indonesia! Kejahatan besar tersebut terjadi baik di bidang sosial ekonomi sumber daya alam hukum tata negara maupun bidang politik ujar Ubedilah Badrun |

| | |
|-----------------|---|
| | <a 1303795896610467841"="" 1303795896610467841\";https:="" edikeceput="" href="http://share.babe.news/s/SSbTkxpQvR;;;\" status="" twitter.com="">http://share.babe.news/s/SSbTkxpQvR;;;\"1303795896610467841\";https://twitter.com/EdiKeceput/status/1303795896610467841 |
| indonesiaunny | 3. Sport Industry memberdayakan ekonomi di tengah krisis akibat pandemi melalui produk dalam negeri di bidang olahraga yang berkualitas dan mempromosikannya di manca negara. @ RafikaBayu @ L0v3D0lph1n @ arimanik @ Jelajah86_ @ anakkosaann @ Magnoolia02pic.twitter.com/CVmGBfXzcD |
| anakBetawinih | Selamat Hari Olah Raga Nasional 9 September 2020 Tema Hari Olah Raga Nasional 2020 ialah Peningkatan Kebugaran Masyarakat dan Prestasi Olahraga Melalui Dukungan Sport Science serta Pemberdayaan Ekonomi Melalui Sport Tourism dan Sport Industry." @ arimanik @ L0v3D0lph1n pic.twitter.com/OogRIN5N6K";;@ @;;\"1303791101195755523\";https://twitter.com/PETJ2020/status/1303791101195755523 |
| PetroekvanDjawa | Tapi sekarang gini deh. Kemaren udah di coba new normal sesuai Sama yang masyarakat budiman mau. Biar aja gapapa kena seleksi Alam yang kuat hidup yang lemah mati. Alasannya? Agar ekonomi tetap jalan kak. Hasilnya? Indonesia dilockdown 59 negara+zona hitam dimana2-- https://twitter.com/squidwardfess/status/1303748010409558016 â€\"1303789376187498497\";https://twitter.com/clapurhandgirl/status/1303789376187498497 |
| wicaksena | klo Anies jdi presiden die jga pasti ambigu buat ambil keputusan itu rakyat Indonesia kan ribet ekonomi merosot marah klo banyak orang sakit kek pandemi ini sewot juga padahal ekonomi sama kesehatan 2 hal yg saling berkaitan dimukabumi\";;\"1303789038277545984\";https://twitter.com/irfanputra13/status/1303789038277545984 |
| cnbcindonesia | Mungkin yang dikehendaki populasi menyusut kue ekonomi meroket. Diidealkan Indonesia diisi SDM sigap bersaing dengan algoritma di masa terdekat. Akumulasi human capital berakselerasi dengan gurita akumulasi primitif. Multitask-human lesser cost. Semoga saya salah. https://twitter.com/BuruhSiluman/status/1303784700549435409 â€\"pic.twitter.com/ZUjEbDikvL\";;\"1303788382435254272\";https://twitter.com/aokhi/status/1303788382435254272 |

2. Contoh Stopword

| username | text |
|-----------------|---|
| suararaa | ekonomi kacau dan pandemi corona terkendali pengamat jokowi maruf amin layak memimpin indonesia politiktoday |
| tempatocoh | tuh orang anggap enteng corona corona dateng joget joget sombong nggk corona berkembang dicuaca panas hancur ekonomi hancur corona berkembang pesat indonesia hebat |
| KontanNews | corona gak lockdown indonesia nekat datang wisatawan asing pekerja asing psbb serahkan daerah ekonomi nyungsep corona tembus utang negara numpuk acaman resesi didepan mata yg salah anies baswedan hahahahhahahahahh |
| clapurhandgirl | anis maya korban corona indonesia jgan rebu ekonomi bangkit lgi |
| aokhi | terapkan psbb lockdown penerapan yg menekan penurunan corona indonesia fokuskan bidang kesehatan dan bidang ekonomi menghindari resesi parah |
| jpnncom | dibilangin video indonesia new rekor baru corona krisis ekonomi menyeret indonesia |
| SyahriAlghifari | negara banned indonesia dan bertambah ekonomi resesi loker seret corona gila |
| indonesiaunny | kesalahan indonesia berpusat jakarta bahkan ekonomi indonesia ditopang dari jakarta jakarta episentrum virus corona yg mengakibatkan lumpuhnya aktivitas ekonomi hancurlah indonesia |
| anakBetawinih | negara banned indonesia dan bertambah lapangan kerja ambyarr ekonomi ancur corona gila prepare worst scenario kawan smua |
| PetroekvanDjava | pemberlakuan jam malam bogor menghancurkan ekonomi jalan pemerintah indonesia cepat memproduksi vaksin ngga nunggu thn karna virus corona berbahaya dari penyakit hiv aids |
| wicaksena | indonesia dari krisis pandemi corona covid krisis ekonomi krisis keamanan krisis korupsi kolusi nepotisme krisis pendidikan krisis |

| | |
|----------------|--|
| | hukum krisis politik perthobatan nasional dan menghargai serikat buruh pekerja pic twitter com ktavvzaynv |
| cnbcindonesia | coronaresesi virus komunis china covid landa dunia pemerintahan joko widodo ngeyel bakal dijamah wabah negeri sangar covid dunia ngakui sdg dilanda resesi ekonomi indonesia ngeyel resesi dng argumentasi |
| AJhunaidhy | corona merebak indonesia idi hak utk bersikap menuntut jokowi ga dgn kesehatan pandemi idi nampak sabarr bgt hasilnya sampe kmrn lho jokowi msh bacotin pemulihan ekonomi mulu |
| dlselfk | coronaresesi virus komunis china covid landa dunia pemerintahan joko widodo ngeyel bakal dijamah wabah negeri sangar covid dunia ngakui sdg dilanda resesi ekonomi indonesia ngeyel resesi dng argumentasi |
| BuruhBuku | corona cepetan ilang biar bisa hidup liburan kota nonton mall majlis dimana ekonomi indonesia pulih ga banyak yg phk sedih bgt |
| ipunglombokGo | dan orang menganggap new marah dianggap fanatik ekstremis bahkan gila corona indonesia tingginya tuntutan ekonomi alasan bahkan kadang lalaikan protokol |
| istridoyi | corona mewabah indonesia pemerintah mengutamakan menyelamatkan ekonomi dibandingkan mencegah penyebaran corona dg lock dll hasilnya geliat ekonomi kunjung korban corona bertambah banyak bahkan kantor pemerintahan |
| bukansopojarwo | tangan kowi rekor indonesia meroket dari ekonomi sampe corona kamiperdulirakyat kamiperdulirakyat |

3. Contoh data bersih dan pelabelan SentiStrength

| Text | Kelas |
|---|----------|
| ekonomi kacau [-4] dan pandemi corona terkendali [4] pengamat jokowi maruf amin layak [4] memimpin indonesia politiktoday , sentimen | neutral |
| tuh [-2] orang anggap enteng corona corona datang joget joget sombong [-4] nggk corona berkembang dicuaca panas hancur [-4] ekonomi hancur [-4] corona berkembang pesat [3] indonesia hebat [3] , sentimen | negative |
| corona gak lockdown indonesia nekat [-2] datang wisatawan asing [-4] pekerja asing [-4] psbb serahkan daerah ekonomi nyungsep [-5] corona tembus utang [-4] negara numpuk acaman resesi didepan mata yg salah [-4] anies baswedan hahahahahahahahahh , sentimen | negative |
| anis maya korban [-3] corona indonesia jgan rebu ekonomi bangkit [3] lgi , sentimen | neutral |
| terapkan psbb lockdown penerapan yg menekan penurunan [-3] corona indonesia fokuskan bidang kesehatan [4] dan bidang ekonomi menghindari [-1] resesi parah [-5] , sentimen | negative |
| dibilangin video indonesia new rekor baru corona krisis [-3] ekonomi menyeret [-4] indonesia , sentimen | negative |
| negara banned indonesia dan bertambah ekonomi resesi loker seret corona gila [5] , sentimen | positive |
| kesalahan [-3] indonesia berpusat jakarta bahkan ekonomi indonesia ditopang dari jakarta jakarta episentrum virus corona yg mengakibatkan lumpuhnya aktivitas ekonomi hancurlah indonesia , sentimen | negative |
| negara banned indonesia dan bertambah lapangan kerja ambyarr ekonomi ancur [-4] corona gila [5] prepare worst scenario kawan smua , sentimen | positive |
| pemberlakuan jam malam bogor menghancurkan ekonomi jalan pemerintah indonesia cepat [4] memproduksi vaksin ngga nunggu thn karna virus corona berbahaya [-4] dari penyakit [-4] hiv aids , sentimen | neutral |
| indonesia dari krisis [-3] pandemi corona convid krisis [-3] ekonomi krisis [-3] keamanan krisis [-3] korupsi [-5] kolusi [-5] nepotisme krisis [-3] pendidikan krisis [-3] hukum krisis [-3] politik perthobatan nasional dan menghargai [5] serikat buruh pekerja pic twitter com ktavvzaynv , sentimen | neutral |
| coronaresesi virus komunis china covid landa dunia pemerintahan joko widodo ngeyel bakal dijamah wabah [-4] negeri sangar covid dunia ngakui sdg dilanda [-3] resesi ekonomi indonesia ngeyel resesi dng argumentasi , sentimen | negative |

| | |
|--|----------|
| corona merebak indonesia idi hak utk bersikap menuntut [-3] jokowi ga dgn kesehatan [4] pandemi idi nampak sabar [4] bgt hasilnya sampe kmrn lho jokowi msh bacotin pemulihan ekonomi mulu , sentimen | positive |
| coronaresesi virus komunis china covid landa dunia pemerintahan joko widodo ngeyel bakal dijamah wabah [-4] negeri sangar covid dunia ngakui sdg dilanda [-3] resesi ekonomi indonesia ngeyel resesi dng argumentasi , sentimen | negative |
| corona cepetan ilang biar bisa hidup liburan kota nonton mall majlis dimana ekonomi indonesia pulih [3] ga banyak yg phk [-4] sedih [1] bgt , sentimen | negative |
| dan orang menganggap new marah [-4] dianggap fanatik [-4] ekstremis bahkan gila [5] corona indonesia tingginya tuntutan ekonomi alasan [-2] bahkan kadang lalaikan protokol , sentimen | positive |
| corona mewabah indonesia pemerintah mengutamakan menyelamatkan ekonomi dibandingkan mencegah penyebaran corona dg lock dll hasilnya geliat ekonomi kunjung korban [-3] corona bertambah banyak bahkan kantor pemerintahan , sentimen | negative |
| tangan kowi rekor indonesia meroket dari ekonomi sampe corona kamiperdulirakyat kamiperdulirakyat , sentimen | neutral |
| dampak corona ekonomi indonesia kukertadaringuinbanten kukertamoderasiberagamapertama kukertaamanproduktifmasacovid pic twitter com ijryzex , sentimen | neutral |



4. Kutipan kamus SentiStrength

| Kata | Bobot |
|------------|-------|
| Woi | 1 |
| Sialan | -5 |
| Jangan | -4 |
| Serius | 2 |
| Baik | 4 |
| Korupsi | -5 |
| Pencuri | -5 |
| Rugi | -5 |
| Membantu | 5 |
| Sumbang | 4 |
| hilang | -4 |
| Kemari | 1 |
| Korban | -4 |
| Kerugian | -5 |
| Keuntungan | 5 |
| Sabar | 2 |
| Menunda | -3 |
| Bertanya | -1 |
| Memperoleh | -2 |

5. Code Program GetOldTweet Python (Grab Data)

```
import sys
if sys.version_info[0] < 3:
    import got
else:
    import got3 as got
def main():
    def printTweet(descr, t):
        print(descr)
        print("Username: %s" % t.username)
        print("Retweets: %d" % t.retweets)
        print("Text: %s" % t.text)
        print("Mentions: %s" % t.mentions)
        print("Hashtags: %s\n" % t.hashtags)
    # Example 1 - Get tweets by username
    tweetCriteria =
    got.manager.TweetCriteria().setUsername('barackobama').setMaxTweets(1)
    tweet =
    got.manager.TweetManager.getTweets(tweetCriteria)[0]
    printTweet("### Example 1 - Get tweets by username
    [barackobama]", tweet)
    # Example 2 - Get tweets by query search
    tweetCriteria =
```

```
got.manager.TweetCriteria().setQuerySearch('europe
refugees').setSince("2015-05-01").setUntil("2015-09-30").setMaxTweets(1)
tweet =
got.manager.TweetManager.getTweets(tweetCriteria)[0]
printTweet("### Example 2 - Get tweets by query search
[europe refugees]", tweet)
# Example 3 - Get tweets by username and bound dates
tweetCriteria =
got.manager.TweetCriteria().setUsername("barackobama").setSince("2015-09-
10").setUntil("2015-09-12").setMaxTweets(1)
tweet =
got.manager.TweetManager.getTweets(tweetCriteria)[0]
printTweet("### Example 3 - Get tweets by username and
bound dates [barackobama, '2015-09-10', '2015-09-12']", tweet)
if __name__ == '__main__':
main()
```



6. Code program cleansing

```
setwd("D:/")
mydata <- read.csv(file.choose(), header = T)
class(mydata)
library("NLP")
library("tm")
df <- do.call("rbind", lapply(mydata$Status, as.data.frame))
df <- sapply(df, function(row) iconv(row, "latin1", "ASCII",
sub=""))
#menghapus url
df=gsub("http[[:alnum:][:punct:]]*", "",df)
#menghapus url
df=gsub("www[[:alnum:][:punct:]]*", "",df)
#menghapus angka kombinasi huruf
df=gsub("[^0-9A-Za-z//' ]", " ",df)
#menghapus angka
df=gsub("[0-9]", " ",df)
df=gsub("[^[:alnum:]]", " ",df)
df <- stripWhitespace(df)
write.table(df, file="cleansing1.csv", sep = ",", row.names = F)
getwd()
```

7. Code Program Stopword

```
package com.uttesh.exude.stemming;
import java.nio.file.*;
import com.uttesh.exude.ExudeData;
import com.uttesh.exude.exception.InvalidDataException;
import static com.uttesh.exude.stemming.Stemmer.c;
import static com.uttesh.exude.stemming.Stemmer.path;
import static com.uttesh.exude.stemming.Stemmer.url;
import java.io.File;
import java.io.FileWriter;
import java.io.IOException;
import java.nio.file.Files;
import java.nio.file.Paths;
import java.sql.DriverManager;
import java.sql.PreparedStatement;
import java.sql.ResultSet;
import java.sql.Statement;
public class Stopwords {
private static String readFileAsString(String fileName) throws
Exception{
    String data = "";
    data = new String(Files.readAllBytes(Paths.get(fileName)));
    return data;
}
```

```

    }

    public static void main(String[] args) throws InvalidDataException,
    Exception {
        try {
            String inputData =
            readFileAsString("/Users/ids/Documents/SKRIPSI/otw
            stopword txt/new folder 2/.txt");

            File newTextFile = new
            File("/Users/ids/Documents/netbins/stopword2/newfo
            lder/.txt");
            FileWriter fw = new FileWriter(newTextFile);
            String output =
            ExudeData.getInstance().filterStoppingsKeepDuplica
            tes(inputData);
            fw.write(output+"\n");
            fw.close();
        } catch (IOException iox) {
            //do stuff with exception
            iox.printStackTrace();
        }
        //System.out.println("output : "+output2);
    }
}

```

8. Code Program Steaming

```
import java.io.BufferedReader;
import java.io.File;
import java.io.FileWriter;
import jsastrawi.cli.output.Output;
import jsastrawi.cli.output.BufferedOutput;
import java.io.IOException;
import java.io.InputStream;
import java.io.InputStreamReader;
import java.sql.Connection;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.HashSet;
import java.util.List;
import java.util.Set;
import java.util.StringTokenizer;
import jdk.nashorn.internal.parser.TokenStream;
import jsastrawi.morphology.DefaultLemmatizer;
import jsastrawi.morphology.Lemmatizer;
import jsastrawi.tokenization.HeuristicTokenizer;
import java.io.IOException;
import java.sql.DriverManager;
import java.sql.PreparedStatement;
```



```

import java.sql.ResultSet;
import java.sql.Statement;
public class Main {
private static HeuristicTokenizer tokenizer;
public static Connection c;

    public static final String driver = "sun.jdbc.odbc.JdbcOdbcDriver";

    public static String url ="jdbc:ucanaccess://";
    public static final String
path="D:"+File.separator+"dbGetDataTwitter.mdb";
    public void setUp() {
        tokenizer = new HeuristicTokenizer();

    }
    public static void main(String[] args) throws IOException {
        Set<String> dictionary = new HashSet<String>();
sendiri
        InputStream in = Lemmatizer.class.getResourceAsStream("/root-words.txt");
        BufferedReader br = new BufferedReader(new InputStreamReader(in));

        String line;
        while ((line = br.readLine()) != null) {
            dictionary.add(line);
        }
        Lemmatizer lemmatizer = new DefaultLemmatizer(dictionary);
        String sql = "SELECT * FROM StopwordsAnies";
        System.out.println("Sedang diproses....");
        try
        {
            c=DriverManager.getConnection(url+path);
            System.out.println("Berhasil konekk database");

```



```

Statement state = c.createStatement();
ResultSet rs = state.executeQuery(sql);
if(rs!=null)
{
while(rs.next())
{
int id = rs.getInt("ID");
String username= rs.getString("username");
String inputData = rs.getString("Status");
String tanggal = rs.getString("tanggal");
int retweet = rs.getInt("retweet");
int favorites=rs.getInt("favorites");
String mention = rs.getString("mentions");
String hastags =rs.getString("hastags");
String geo_location=rs.getString("geo_location");
StringTokenizer st = new StringTokenizer(inputData);
System.out.println("---- Split by space -----");
StringBuilder sb = new StringBuilder();
String temp="";
while (st.hasMoreElements()) {
temp = temp +
lemmatizer.lemmatize(st.nextToken())+" ";
//sb.append(
lemmatizer.lemmatize(st.nextToken())+" ");
}
//String sql2="UPDATE Stopwords set Status = ? where
ID=?"; //versi edit
String sql2 = "INSERT into StemmingAnies
values(?,?,?,?,?,?,?,?)"; //versi insert
System.out.println("output ID "+id+" : "+temp);
PreparedStatement pStmt =

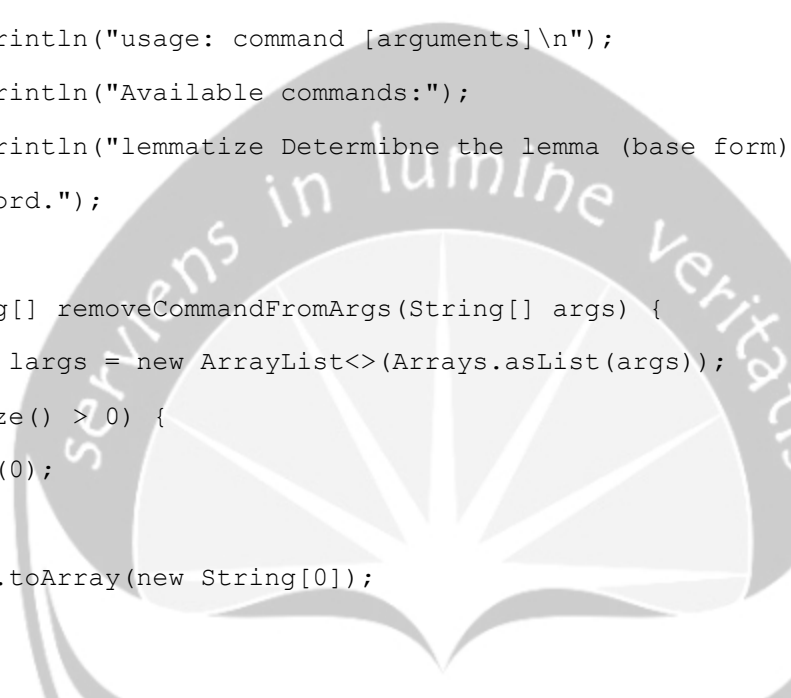
```

```

c.prepareStatement(sql2);
// pstmt.setInt(1, c.GetRowDataTwitter());
pstmt.setInt(1, id);
pstmt.setString(2, username);
pstmt.setString(3, temp);
pstmt.setString(4, tanggal);
pstmt.setInt(5, retweet);
pstmt.setInt(6, favorites);
pstmt.setString(7, mention);
pstmt.setString(8, hastags);
pstmt.setString(9, geo_location);
pstmt.executeUpdate();
}
}
rs.close();
state.close();
c.close();
System.out.println("database ditutup");
}
catch(Exception EX)
{
System.out.println("Error Reading From database. . .");
System.out.println(EX);
}
if (args.length == 0) {
printHelp();
} else if (args[0].toLowerCase().equals("lemmatize")) {
Output bufferedOutput = new BufferedOutput();
LemmatizeCmd lemmatizeCmd = new LemmatizeCmd(bufferedOutput);
lemmatizeCmd.handle(removeCommandFromArgs(args));
System.out.print(bufferedOutput.toString());
}

```

```
} else {  
    printHelp();  
}  
  
private static void printHelp() {  
    System.out.println("JSastrawi");  
    System.out.println("usage: command [arguments]\n");  
    System.out.println("Available commands:");  
    System.out.println("lemmatize Determinbne the lemma (base form)  
for a given word.");  
}  
  
static String[] removeCommandFromArgs(String[] args) {  
    List<String> largs = new ArrayList<>(Arrays.asList(args));  
    if (largs.size() > 0) {  
        largs.remove(0);  
    }  
    return largs.toArray(new String[0]);  
}  
}
```



9. Hasil akurasi metode Naïve Bayes Classifier menggunakan K-10 Folds Cross Validation

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      103166           80.3617 %
Incorrectly Classified Instances    25211           19.6383 %
Kappa statistic                    0.7052
Mean absolute error                 0.2183
Root mean squared error             0.3067
Relative absolute error             49.27 %
Root relative squared error         65.1562 %
Total Number of Instances          128377

=== Detailed Accuracy By Class ===

```

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | C |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|---|
| | 0.823 | 0.109 | 0.763 | 0.823 | 0.792 | 0.700 | 0.959 | 0.917 | |
| | 0.869 | 0.032 | 0.939 | 0.869 | 0.903 | 0.852 | 0.974 | 0.962 | |
| | 0.717 | 0.151 | 0.709 | 0.717 | 0.713 | 0.564 | 0.886 | 0.784 | |
| Weighted Avg. | 0.804 | 0.095 | 0.809 | 0.804 | 0.805 | 0.709 | 0.940 | 0.888 | |

```

=== Confusion Matrix ===

  a    b    c  <-- classified as
31579   88 6717 |   a =  positive
  9 40457 6084 |   b =  negative
9773  2540 31130 |   c =   neutral

```

10. Hasil akurasi metode Support Vector Machine menggunakan K-10 Folds Cross Validation

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      128217          99.8754 %
Incorrectly Classified Instances      160          0.1246 %
Gamma statistic                     0.9981
Mean absolute error                  0.0008
Root mean squared error              0.0288
Relative absolute error              0.1875 %
Root relative squared error          6.1242 %
Total Number of Instances           128377

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Cl:
0.997   0.000   1.000   0.997   0.999   0.998   0.999   0.998   1
0.999   0.000   1.000   0.999   0.999   0.999   0.999   0.999   1
1.000   0.002   0.996   1.000   0.998   0.997   0.999   0.996   1
Weighted Avg.   0.999   0.001   0.999   0.999   0.999   0.998   0.999   0.998

=== Confusion Matrix ===

  a    b    c  <-- classified as
38284    0   100 |    a =  positive
  0 46490    60 |    b =  negative
  0     0 43443 |    c =   neutral

```